Numerical Weather Prediction

The Impact of Observations on Mesoscale Model Forecasts of 3-hourly Rainfall Accumulations - Updated Results

Forecasting Research Technical Report No. 385

Simon Anderson

email: nwp_publications@metoffice.com

©Crown Copyright

The Impact of Observations on Mesoscale Model Forecasts of 3-hourly Rainfall Accumulations – Updated Results

S.R. Anderson

Contents

Section 1	Introduction	Page 2
Section 2	Case selection	Page 2
Section 3	Method for assessing impact of observations	Page 3
Section 4	Example case study	Page 6
Section 5	Results from all the cases	Page 9
Section 6	Discussion of results	Page 11
Section 7	Conclusions and recommendations	Page 15
Appendix 1	Further details of case selection technique	Page 18
Appendix 2	Map showing locations of surface and radiosonde observations	Page 19

The Impact of Observations on Mesoscale Model Forecasts of 3-hourly Rainfall Accumulations – Updated Results

S.R. Anderson

1. Introduction

1.1 Background and aims of study

The primary aim of this work has been to investigate the impact of observations on the accuracy of forecasts of precipitation accumulation from the Met Office's Mesoscale NWP model. Results from such experiments provide useful evidence on which to base the design of observational networks over and around the UK. This study is a follow-up to a project carried out in the mid-1990s and documented in Anderson *et al.* (2000) – hereafter referred to as "Study A". In this study, however, a new model domain and data assimilation scheme have been used and there have been differences made to the method, following recommendations made in the previous report. Precipitation was again chosen as the impact variable because of the continued interest in improving its forecasts. All current observations made within the Mesoscale model domain have been considered. A full description of the 12km resolution, 38-level model (operational since June 1998) can be found in Wilson and Culverwell (1997).

1.2 Overview of this report

The paper is divided into the following sections. Section 2 contains an analysis of the procedure used to select the cases, and gives useful guidance on the frequency with which we might expect observations to deliver marked forecast benefits. The case study method used to obtain the results is described in section 3. Section 4 contains an example case study.

A summary of all the results is given in section 5. A discussion of the results is offered in section 6, and conclusions and recommendations in section 7.

2. Case selection

2.1 Technique

Cases in which initial data had a marked benefit on forecasts of widespread rainfall were identified by looking for notable improvements in forecast skill between consecutive operational forecasts valid at the same time (following Graham *et al.*, 2000). Improved skill may be attributed to either the additional observations available for assimilation in the later run or the impact from the more recent boundary conditions used in the later run. Cases in which objective measures of forecast skill indicated a marked improvement in

the T+6h forecast over the T+12h forecast, or the T+12h forecast over the T+18h were short-listed for study. Later, the case selection software was altered to short-list improvements in the T+18h forecast over the T+24h, or the T+24h forecast over the T+30h.

More details of the criteria used for the case selection procedure are given in Appendix 1, together with the definitions used for 'marked benefit' and 'widespread'. In all, fifteen cases were identified for detailed study, and it is the results from these that are presented in this paper.

2.2 Frequency of marked data impact on rainfall forecasts

In this sub-section, we provide an estimate of the frequency with which we can expect significant data impacts on forecasts of rainfall accumulation. This calculation is based on results of the case selection procedure used to select our first eleven cases, between the months of March and October 2000. (Our remaining four cases were chosen using a similar procedure, but during the second half of 2001).

Between March and October 2000, a total of 710 operational 00Z, 06Z, 12Z and 18Z Mesoscale model forecasts were checked for sensitivity to initial data. Of these, 585 forecasts (~82%) were for cases in which 'widespread' rain (see Appendix 1) was observed over the UK area at the verifying time. Of these 585 cases, 93 (~16%) showed 'marked' positive impacts in consecutive forecasts, either from recent observations (i.e. observations from within the Mesoscale domain assimilated in the 6-hour period preceding the forecast) or from updated model boundary data.

Subjective criteria, described in Appendix 1, were used to select cases from the 'shortlist' of 93 as a representative sample. Eleven cases were selected. Assuming the cases are representative, we may tentatively conclude that marked improvements occur between consecutive model forecasts in about 16% of all forecasts of widespread rain (corresponding to about 13% of the time).

N.B. Of the 585 cases of 'widespread' rainfall, 13 (~2%) showed 'marked' *degradations* between consecutive forecasts, based on a *reverse* of the objective criteria described in Appendix 1.

3. Method for assessing impact of observations

3.1 Experiment format

To assess the impact of the observations, a series of model re-runs was performed for each case. For this study, all the data impact experiments comprised a 6-hour period of assimilation (using the Met Office 3D-var assimilation scheme - see Macpherson *et al.*, 2002) followed by a forecast.

For each case, two control runs were performed. The ALL OBS run, which used all the available data (i.e. as used in the operational Mesoscale runs), and the NO OBS run which entailed a dummy assimilation using no data. The NO OBS run represents the

earlier, poor forecast, whilst the ALL OBS run represents the next operational model forecast, with all the benefit of the 6 hours of observations made between the two runs.

All our re-run experiments were designed to mimic as closely as possible the operational model runs. However, there was a difference in how we utilised the model boundary data. Operationally, the boundary data were updated every 3 hours, the fields being created from the Met Office Global model. In our re-runs, we used identical boundary conditions for both the NO OBS and ALL OBS runs. This meant that, provided the ALL OBS forecast still showed a marked improvement over the NO OBS forecast, we could conclude that the improvement in the forecast had come from the observations, rather than through the model boundary.

For the remaining experiments in each case study, each observation type was denied from the assimilation individually, and the impact recorded (as described in Sub-section 3.3). This method of assessing the impact of observation types separately provides useful insight into which data types play key rôles in an observation network - a requirement for network design. (N.B. The alternative method of including each observation type individually in the assimilation process was used in Study A – see Anderson *et al.*, 2000. The data-denial technique, employed here, is more realistic to what may happen operationally, if a communications network were to crash for example. There is the disadvantage of data-redundancy, however (i.e. if experiments are run omitting one data type, then it is likely that its absence will be compensated by other remaining types)).

3.2 Observation types used

The observation types currently used in the Met Office Mesoscale model include profile data from radiosonde observations (of wind, temperature and humidity) and wind profilers, aircraft observations (of wind and temperature) and surface observations (of PMSL, wind from SYNOPS, moored buoys and ships, and temperature and humidity from SYNOPS). All these data types are assimilated using the Met Office 3D-var scheme. A map showing the locations of the SYNOP stations, moored buoys, rigs, wind profilers and radiosonde stations whose data were assimilated in the model is given in Appendix 2. (The map also shows the airports currently served by AMDAR-equipped planes.)

In addition to the above, MOPS (Moisture Observation Processing System) data are assimilated. MOPS comprises 3-D cloud cover information and 2-D rainrate information from the Met Office's nowcasting system Nimrod (see Macpherson *et al.*, 1996). Briefly, Nimrod produces analyses from satellite and radar rainfall imagery and selected SYNOP observations combined with a short-period mesoscale forecast. The elements of the SYNOP used are cloud cover, base and type and present weather. The analysis of cloud fraction is used as an observation type in the Met Office analysis correction scheme – see Lorenc *et al.*, 1991) as "pseudo radiosondes" to influence the humidity analysis, with one profile per grid square. Rainrate is assimilated using the same scheme via a Latent Heat Nudging technique (see Jones and Macpherson, 1997).

3.3 Forecast verification

Forecast skill was assessed by comparing the accumulation of precipitation over the final 3 hours of the forecast with an analysis of 3-hour accumulation derived from the UK weather radar network by Nimrod. The Nimrod analyses include corrections for range and bright band effects, and for orographic enhancement of precipitation (Kitchen *et al.*, 1994). Before use for forecast verification, the Nimrod analyses, which have a grid resolution of 5km, were smoothed by assigning the average value over a square array of 7x7 grid points to the central point in the array. The forecast accumulations were smoothed in a similar way using a 3x3 grid-point array. The smoothing yielded observed and forecast precipitation patterns that had similar scale representation (~36km). Verification of the model forecasts was restricted to the area covered by the UK weather radars (Fig. 1).

Equitable Threat Scores (ETS) were calculated at two lower thresholds of rainfall accumulation: 0.5mm/3h and 2.0mm/3h. These rates were based on the thresholds for verification applied operationally at the Met. Office (namely 1.0mm/6h and 4.0mm/6h). The ETS (see Schaefer, 1990, for further details) is widely respected, and is used as the main verification measure for precipitation in the Met Office's NWP Index. It penalises both missed forecasts of observed rain and the forecasting of rain in the wrong place, and includes a correction for the rainfall events which may have been expected to be correctly forecast by chance. It has a theoretical range from 1.0 (highest skill) to -0.33.

The skill of each forecast is expressed as a percentage of the difference in skill between the ALL OBS and NO OBS controls. For example, for a forecast using data from MOPS,

$$\text{BENEFIT OF MOPS} = \frac{\text{ETS(MOPS)} - \text{ETS(NO OBS)}}{\text{ETS(ALL OBS)} - \text{ETS(NO OBS)}} \times 100\%. \qquad (1)$$

Hit rates, false alarm rates and root mean square factor (rmsf) errors are other measures commonly applied to precipitation verification. They were also used, therefore, to assess the skill of the rainfall forecasts at the lower threshold of 0.5mm/3h.

The hit rate represents the percentage of points where rainfall accumulations above the threshold were forecast correctly. Hence, high scores are preferred. The false alarm rate represents the percentage of points where rainfall accumulations were forecast but did not occur. Low scores are desirable.

The root mean square factor error can be thought of as an average ratio, in a root-mean-square sense, of the forecast to observed precipitation. Low scores are once again preferred. It has the advantage of concentrating on the rain areas rather than being dominated by dry events, as, for every grid-point where the score is calculated, at least one of the forecast value or observation has to exceed the threshold value. (Note that the threshold value for the rmsf error is used as a lower limit in order to avoid division by zero). The main disadvantage, however, is that it condenses all the information about the forecast behaviour at different rainfall intensities into one score. Hence it is considered

inferior to the ETS, which may be calculated at any number of rainfall thresholds.

Finally, the bias was analysed by calculated the frequency bias index, which is the number of forecast rain points divided by the total number observed.

4. Example case study

In this section we describe one of the case studies (Case 11: verifying at 06UTC, 7th October 2000) to demonstrate the procedure for calculating the impact of the observation types and to give the reader a better understanding of the magnitudes of the skill scores that contribute towards the results from all fifteen cases. The method of objective verification used to obtain the results is presented, and we also give a subjective evaluation of the observational impact on forecast skill for both low and high levels of rainfall accumulation (i.e. for thresholds of 0.5mm/3h and 2.0mm/3h).

4.1 Synoptic situation and precipitation distribution

The surface analysis for 06UTC on 7th October 2000 is shown in Fig. 2a. A high pressure system was located to the south west of the British Isles, with a fairly strong zonal flow between 50˚ and 60˚N to its north. The zonal flow contained a succession of frontal systems, which slowed up over the UK as they approached a ridge of high pressure over central Europe. The significant weather chart for 06UTC (Fig. 2b) shows widespread precipitation associated with the slow-moving fronts, affecting most of central England.

Fig. 3 shows the contoured precipitation accumulation field for the three hours 03-06UTC. The main areas of accumulations in excess of 2.0mm are over NW England, the Midlands and central southern England, with accumulations peaking at over 8.0mm. The Nimrod analysis is clearly consistent with the significant weather reports (Fig. 2b).

4.2 Control runs

The corresponding 6-hour forecasts from the NO OBS and ALL OBS control runs are shown in Figs. 4a&b respectively. Both forecasts produce areas of rainfall positioned N-S over the UK. The ALL OBS forecast, however, is further advanced and predicts accumulations peaking at over 8.0mm. Comparison with the analysis (Fig. 3) shows clearly that the ALL OBS run verifies better, providing useful guidance, whilst the NO OBS run is very misleading.

The statistics for the two control runs for this case are shown in Table 1. The superiority of the ALL OBS run is reflected in a higher hit rate at the 0.5mm/3h threshold (65.2% compared to 60.7% for the NO OBS run), a lower false alarm rate and rmsf error, and higher ETS scores at both thresholds (0.276 compared to only 0.054, for example, at the 2.0mm/3h threshold). The greater skill of the ALL OBS run shows that observational data introduced in the 6-hour assimilation cycle prior to the ALL OBS forecast delivered a marked benefit.

	Threshold	E.T.S.	Hit Rate	False Alarm Rate	RMS Factor Error
NO OBS	0.5:	0.201	60.7	50.5	4.41
	2.0:	0.054			
ALL OBS	0.5:	0.306	65.2	39.5	3.48
	2.0:	0.276			

Table 1 - Scores for Case 11 from different verification measures for precipitation forecasts at T+6h for the control forecasts.

This table shows the results for Case 11 of the two control runs.
Column 2 indicates which rainfall threshold the scores refer to - either the lower threshold of 0.5mm/3h, or the higher one of 2.0mm/3h.
For an explanation of the verification measures in columns 3-6, see sub-section 3.3.

4.3 Impact of individual observation types

We now investigate the rôle that the individual observation types played in the impact. The forecasts were re-run, and each time a different observation type was individually denied from the assimilation, and the effect on the calculated forecast skill scores noted. Table 2 shows the results of these runs for Case 11.

The ETS results show that the individual observation type that caused the largest degradation when excluded, for this case, was the moisture field created by MOPS. When denied from the assimilation, the score of 0.252 at the lower threshold represents a 51% degradation compared to the score using NO OBS, when calculated using Equation (1). The other observation types that gave marked degradations at the lower threshold when removed were radiosonde data (RsWTH, 19%) and aircraft observations (17%). At the higher threshold, the two largest degradations were caused by denying MOPS (43%) and radiosonde data (40%).

For Case 11, two or more of the verification measures show the importance of MOPS, radiosonde observations and/or aircraft data.

Fig. 4c shows the forecast denying the assimilation of MOPS data. The peak forecast rainfall totals have been much reduced and the frontal rain band is less far advanced. The forecast denying only radiosonde observations (not shown) predicts similar rainfall totals, but the frontal rainband positioning is very similar to the ALL OBS run. It can be implied that although the radiosonde observations help to produce a better forecast of the rainfall amounts, it is only the MOPS information that correctly advances the rain far enough east.

	Threshold	E.T.S.	Hit Rate	False Alarm Rate	RMS Factor Error
ALL OBS -	0.5:	0.252	57.8	42.8	3.92
MOPS	2.0:	0.181			
ALL OBS -	0.5:	0.332	70.4	39.1	3.22
RsH	2.0:	0.306			
ALL OBS -	0.5:	0.286	73.0	45.1	3.46
RsWTH	2.0:	0.188			
ALL OBS -	0.5:	0.322	68.9	39.4	3.44
SURFACE	2.0:	0.283			
ALL OBS -	0.5:	0.288	61.7	40.0	3.63
AcWT	2.0:	0.276			
NO OBS -	0.5:	0.215	60.9	48.9	4.52
Boundary data	2.0:	0.056			

Table 2 - Scores for Case 11 from different verification measures for precipitation forecasts at T+6h for the model runs denying individual observation types.

This table shows the results for Case 11 from the model runs denying observation types individually. The thresholds and verification measures are the same as in Table 1.
The observation types tested were MOPS (see Sub-section 3.2 for details), radiosonde humidities (RsH), radiosonde winds, temperatures & humidities and wind profiler data (RsWTH), surface observations, aircraft winds and temperatures (AcWT) and the model's lateral boundary data.
All scores should be compared to the control run scores shown in Table 1.

4.4 Impact of combined observation types

Further runs were carried out denying combinations of observation types in order to find a small subset of types that would produce a forecast closely resembling the NO OBS control run. For Case 11, a run denying both MOPS and radiosonde observations showed how denying only two observation types could almost match the effect of using

no observations whatsoever in the assimilation. The results from this run are given in Table 3.

ALL OBS -	Threshold	E.T.S.	Hit Rate	False Alarm Rate	RMS Factor Error
MOPS	0.5:	0.199	60.8	50.8	4.21
& RsWTH	2.0:	0.048			

Table 3 - Scores for Case 11 from different verification measures for precipitation forecasts at T+6h for a model run denying a combination of observation types.

This table shows the results for Case 11 from a model run denying a combination of observation types. The thresholds and verification measures are the same as in Table 1.
Note how all the skill scores show results for the combination run denying MOPS & RsWTH comparable to the NO OBS scores in Table 1.

Fig. 4d shows the forecast from the run denying MOPS and radiosonde observations. The chart for this run compares very well with the NO OBS forecast (Fig. 4a) over virtually the whole of the Nimrod domain. The ETS has dropped to 0.048, slightly lower than the NO OBS value. The example highlights the data redundancy that can occur (i.e. in this case the redundancy of the surface and aircraft data if other observation types are available).

5. Results from all the cases

Results like those obtained in Case 11 were computed for each case. Out of the fifteen cases studied, thirteen showed impacts from observations within the Mesoscale model domain. The remaining two cases showed strong impacts from the model lateral boundary data, and will be presented later.

5.1 Observational impacts

The results from the thirteen cases showing observational impacts have been combined to produce the information in Figure 5. The upper histogram displays the results for ETS at the lower threshold, whilst the lower histogram represents the higher threshold rainfall results. The vertical scale represents the number of times that each individual observation type gave a noticeable degradation to the model forecast when it was denied from the assimilation. In more detail, each bar shows the number of times that the degradation exceeded 25% of the NO OBS score (i.e. when no observations were used), and also when the degradation exceeded 50%. Results for the eleven shorter-range T+6h and T+12h forecasts and results from two longer-range T+18h and

T+24h forecasts have been combined.

The results from thirteen case studies, using ETS as the measure of skill, are summarised as follows:

- At the lower threshold of 0.5mm/3h, the most noticeable degradations to the rainfall accumulation forecasts were caused when either the **MOPS analyses**, **surface observations** or **aircraft data** were denied from the assimilation. The profile information from **radiosondes** played a slightly more supporting rôle.

- Similar results were gained at the higher threshold of 2.0mm/3h.

- Few examples of marked impact of radiosonde humidity data were found.

The bias scores were also analysed. The effect of denying each observation type individually usually had only a small effect on the bias score. The only trends to appear were that when surface observations were omitted, the forecasts tended to under-predict the rainfall at both accumulation thresholds (when compared to the bias scores from the All Obs runs), and when radiosonde data were omitted, over-prediction tended to occur.

As mentioned in Section 3, other objective skill scores (all at the lower rainfall threshold of 0.5mm/3h) were calculated. A brief summary of these results is as follows.

- Over all the cases, the largest impacts to the hit rate scores were seen when **surface observations**, **aircraft data** or the **MOPS analyses** were denied from the assimilation.

- When considering the effect on scores of false alarm rate, the largest degradations were caused by the removal of **radiosonde data**, **surface observations** or the **MOPS analyses**.

- For the root mean square factor error, scores were degraded the most by denying the **MOPS analyses**, **surface observations** or **aircraft data**.

These results largely re-enforce those found using the ETS skill score.

5.2 Boundary data impacts

All the cases were re-run to test the effects of the model lateral boundary conditions. This was done by running the No Obs forecasts using non-updated boundaries, and comparing results with the No Obs Control run (which used the same updated boundary conditions as the All Obs Control).

For the shorter range (T+6h and T+12h) forecasts, the statistical results showed that, in the majority of cases, the forecasts were very similar to the No Obs control runs.

These results presented as scatterplots (see Fig. 6) also show a distinct lack of spread (even for the second plot where the scale has been enlarged). Comparing the rainfall forecasts by eye showed that in each case these two runs gave almost identical results (even for the statistical outlier cases, Cases 5, 7 and 12, where the updates in boundary conditions were shown to have slight negative effects statistically).

Experiments were run to test how far the model boundary data influence could extend into the Mesoscale model domain during a 12-hour forecast (as followed using a tracer variable). Three cases were studied - Case 5 (where the updated boundary data gave a slight negative impact), 10 (a slight positive impact) and 12 (the best example of a mobile situation). An example (from Case 12) showed that the information from the model's western boundary extended as far as central Britain at ~500 hPa and eastern Britain at ~200 hPa within 12 hours.

In spite of this, the evidence from the first eleven cases all suggested that the impact from the model boundary data on these short range (T+6h and T+12h) forecasts was very small. The result is perhaps not surprising, as both the short forecast lengths and the less mobile summer synoptic regimes may have meant that the influence of the boundary data was felt very little. It was therefore decided to change the case selection criteria, to look at some longer range (T+18h and T+24h) cases during more mobile autumn regimes.

Of the four longer range cases studied, Case 18 and Case 19 both showed boundary data impacts (i.e. benefits from observations outside the Mesoscale model domain, interpolated from the Global model grid). In particular, the later (T+24h) forecast in Case 18 has been shown to be improved *entirely* by the updated boundary conditions (with negligible improvement coming from the assimilated observational data).

The remaining two cases both showed observational data impacts (from MOPS and radiosondes). These results were combined with the results from the shorter-range studies for the histograms in Fig 5.

6. Discussion of results

(i) Which observations are currently working well in the Mesoscale model assimilation?

Looking at the results (not presented) from the individual data denial re-runs from all the case studies, the implication is that in most cases, the denial of data from individual observation types leads to quite minor degradations to the forecasts. This would imply a certain degree of redundancy in the observational data. The results displayed in histogram form in Fig. 5 show that each observation type recovered more than 50% of the NO OBS impact on the forecast in at least one case. Surface, MOPS and aircraft data all had similar impacts in at least three cases, at both rainfall thresholds, with radiosonde data playing a more supporting rôle.

Note that the "25%" and "50%" impact amounts have been used so that the results

can be ‘compared' with those found in Study A in the mid-1990s (see Fig. 7, Anderson *et al.*, 2000). However, in that study an entirely different method of calculating results was used (i.e. data inclusion rather than data denial), along with the previous Mesoscale model domain and AC assimilation scheme. The current results show the continued importance of MOPS and surface data - see **(vi)** below - while radiosonde data now play less of a rôle – see **(iii)** below.

The relative benefit of aircraft data has increased substantially, and is indeed one of the main results of this study. It is worth noting here that in six cases, the benefits to the forecasts were delivered by observations made at 00Z. At this time there are few AMDAR reports, which effectively rules them out from providing a major impact. Out of the remaining seven cases, the aircraft data gave notable benefit in six, thus demonstrating further the influence of aircraft data.

(ii) Which observations are important in cases with heavy precipitation?

Evidence from Fig. 5 suggests that aircraft and surface data gave the most major impacts at the higher rainfall threshold of 2mm/3h (in four cases each), with MOPS and radiosonde data each providing major impacts in one less case.

Studying the cases themselves, there appeared to be nine cases with precipitation of over 8mm in 3 hours, although this was a very subjective estimate and the areas affected were usually only the size of a couple of counties. Two of these cases featured mainly convective precipitation, whilst the rest were frontal. Radiosonde data were beneficial in seven of these nine cases, MOPS in five, surface data in four cases and aircraft data in only one. This result suggests that radiosonde measurements provide important input for modelling the 'dynamical' ascent of air in cases of widespread moderate or heavy rain (e.g. frontal rain with embedded convection). Furthermore, for significant enhancement of rainfall over hills (e.g. as a result of the seeder-feeder process), the wind direction and speed in the lower troposphere (plus of course the humidity) are particularly important.

(iii) Which observation types do not contribute much?

Results from Fig. 5 show that radiosonde data gave fewer major impacts than all other observation types. The impact of removing the sonde data individually is more likely to be modest, based on the cases studied, though in many cases (9 out of 13) when combined with one other denied observation type, the impact was major. This implies that radiosonde data play a strong supporting rôle.

Note that the data denial technique used for this study does not make this question easy to answer. When one observation type is denied from the assimilation, it is quite possible for another data type to compensate for the loss, seemingly showing the redundancy of the observations being tested. This is likely to have been the case when the radiosonde humidity profile data were denied, as discussed in **(vi)** below.

(iv) What is the magnitude of the influence on the forecasts from the boundaries

with the Global model?

All the cases were re-run to test the effects of information in the model's lateral boundaries. The evidence from the eleven short range (T+6h and T+12h) cases all suggested that the impact from the model boundary data on the forecasts was very small. This is probably because of the combination of short forecast ranges and less mobile (summer) synoptic regimes.

Of the four longer range (T+18h and T+24h) cases studied, two showed boundary data impacts. Both of these cases featured strong flow through the model boundary, which may imply that the model boundary data will usually only influence the forecast during mobile synoptic conditions and after a lead-time in excess of 12 hours.

(v) Which observation types make an important contribution to the prediction of precipitation (especially convective) in weak synoptic forcing?

There were no case studies identified during periods of weak synoptic forcing. The reason for this is likely to be because of the 'widespread rainfall' threshold used in the case-selection procedure. In particular, accumulations of over 0.5mm had to have been observed at over 25 grid points, which may have ruled out a number of localised events.

(vi) What is the impact of radar and satellite data – assimilated through MOPS - and surface observations?

Both MOPS and surface data have markedly degraded the forecasts when denied (as displayed in Fig. 5).

MOPS moisture data were shown to rank highly in four of the eleven short-range cases (Cases 7, 9, 10 and 11). This finding is perhaps not surprising with rainfall as the impact variable. Further re-runs were performed to investigate which components of MOPS gave the most impact - the cloud (relative humidity) or the radar precipitation part of the data. Results showed that the *precipitation* component (assimilated using the latent heat nudging technique) was overwhelmingly the most beneficial, apart from in Case 11 where the results for each component were similar. Similar conclusions were found from the longer-range cases (Cases 15 and 16).

MOPS data were also shown to be important by highlighting the redundancy of the radiosonde humidity data when these were denied from the assimilation. This can be seen by the small number of cases where denying sonde humidities caused a major degradation to the forecast. Although the radiosonde data give direct measurements of humidity, the upper-air network (with an average spacing upwards of 200km - see map in Appendix 2) is not dense enough for mesoscale variability to be adequately resolved. An advantage of the MOPS process is that it receives a more varied observational input, leading to a better humidity analysis on a smaller horizontal scale.

Surface observations ranked significantly in over half of all cases studied (Cases 1, 4, 5, 6, 8 and 12). (Interestingly, nearly all these cases were from spring and summer months.) This evidence, together with the findings of Study A (2000), shows that there is a key rôle for surface data, including cloud information (through MOPS) and that the

impact on NWP forecasts should still be considered in any plans to modify the UK surface network. (The map in Appendix 2 shows the locations of the surface stations.)

For the four cases where surface observations showed the most major impacts, re-runs were performed where only surface pressure data (PMSL and station-height pressure) were denied. The results were inconclusive however - in one case (Case 5) the surface pressure accounted for all of the data impact, in one case (Case 4) it accounted for none, and in the remaining two cases (Cases 6 and 8) roughly 40%.

(vii) What is the dependence on large scale vs. convective precipitation, frontal activity, propagation speed, forecast length, and distance from boundary?

Most of the cases involved frontal rainfall. In these cases, *all* observation types have shown benefit to at least one of the forecasts. Out of the two cases of convective rain (Cases 4 and 8), both have shown benefit from surface observations and one from radiosonde data (Case 4) (though in Case 8, radiosonde data made the forecast *worse*).

The main dependencies on forecast length seem to be that MOPS data tended to be of benefit to short-range T+6h forecasts (Cases 7, 9, 10 and 11) and aircraft data tended to benefit the T+12h forecasts (Cases 1, 2, 5 and 12), but the sample is very small. (NB Two longer range cases (Cases 15 and 16) also showed benefits from MOPS data and Case 8 showed a large T+6h aircraft data benefit). Model boundary data were only shown to benefit the longer range T+18h and T+24h forecasts (Cases 18 and 19), where the rain areas had advected a suitable distance through a boundary into the model domain.

Plots have been produced showing timeseries of ETS and ln(bias) scores through each of the forecasts (for the All Obs and No Obs control runs and the 'No Obs/No updated boundary conditions' runs) - see Fig. 7 for an example plot (from Case 10). For the ETS timeseries, the most obvious trend was that the scores for the All Obs control runs were noticeably higher than for the other two runs, throughout the forecasts, but usually with some convergence by T+12h. In more cases than not, the ETS decreased slightly throughout the forecast, though there were undulations in the scores and each curve behaved differently. There were no obvious trends amongst the ln(bias) plots. Most biases were fairly small, and the cases with the largest extremes showed nicely how the use of observations helps to reduce the bias in the forecasts.

(viii) Comparison with previous work

This study most closely compares to Study A by Anderson *et al.* (2000), although changes have been made to the method employed and to both the model and assimilation scheme used. Because of this, the histograms produced (Fig. 5 here and Fig. 7 in the referenced paper) cannot be directly compared. However, some likenesses and trends can be commented on with a degree of certainty:

- The rôles of the MOPS analyses and of surface observations are confirmed again. Note that surface observations were also found to be beneficial in a Global model impact study by Graham *et al.* (2000) and a Limited Area model

impact study (during FASTEX) by Anderson (1998).

- The utility of aircraft observations has increased relatively since the study in the mid-1990s. It is likely that this is due to the vast increase in the number of AMDAR reports of wind and temperature during the take-off and landing of aircraft at most major airports over NW Europe in recent years. Note that aircraft data at cruise level over the N. Atlantic have previously shown benefit in Global model impact studies (e.g. Graham *et al.*, 2000).
- The relative utility of profile observations from radiosondes and wind profilers appears to have decreased, though this may be a symptom of the data denial method employed (as described in **(iii)** above). Alternatively, it could be linked to the general reduction in the UK radiosonde network. Note that although the humidity profile information may appear to have declined the most (since the wind and temperature components have been compensated by the increased use of AMDAR data and wind profiler observations), the MOPS humidity analyses (at every model grid point) seem to compensate.

7. Conclusions and recommendations

7.1 Conclusions

The following conclusions may be drawn from the results of our fifteen case studies looking at the impact of observations over the Met Office Mesoscale model domain (and its lateral boundary conditions) on forecasts of widespread precipitation accumulation over the UK. Note that "Study A" is an earlier work by Anderson *et al.* (2000), referred to throughout this paper:

- Data from all observation types currently assimilated by the Mesoscale model have been shown to be of significant benefit to the short range forecast in at least one case study, recording a degradation of over half of the total observational impact when individually removed from the assimilation.

- The 3-D moisture field produced by the Met Office Moisture Observation Processing System (MOPS - see sub-section 3.2 for details) has, as in Study A, been shown to be highly beneficial when assimilated into the Mesoscale model. In this study, there were significant degradations in eight out of fifteen cases when the MOPS data were removed. Furthermore, it has been shown that the radar precipitation data supply the most beneficial component to the MOPS analyses (over and above the cloud data).

- Aircraft observations of wind and temperature now provide significant benefits in a sizeable proportion of cases (six out of fifteen *). This is in line with the increase in AMDAR reports over NW Europe since Study A was completed in the 1990s. (* Note that this ratio increases to 'six out of nine', if we discount the cases where AMDAR

reports are limited due to night-flight restrictions.)

- Surface observations have shown marked benefits to the forecasts of rainfall accumulations (in five cases).

- Radiosonde profiles of wind, temperature and humidity (and data from the wind profiler network) now provide more of a supporting rôle than in Study A. This drop in significance is in line with the (continued) decline in the radiosonde network in recent years. The decrease in humidity profile information appears to be compensated by the use of MOPS analyses.

- It is estimated tentatively that observations made within the Mesoscale model domain deliver marked benefits over and above the model background field in about 16% of Mesoscale model forecasts of widespread rainfall up to T+12h (using the definitions in Appendix 1 for 'marked benefit' and widespread').

- At shorter lead times (up to T+12h), when marked improvements occur between consecutive model forecasts, it is observational input, rather than updated model boundary information, which is usually responsible. Model boundary data are more likely to become important at longer forecast ranges (e.g. T+18h and over) and in more mobile weather situations.

7.2 Recommendations

It is recognised that the results from just fifteen case studies may not fully demonstrate the true significance that each observation type makes over a longer period. However, they generally reinforce the findings from Study A. The changes in impact from given observation types appear to be consistent with differences in the volumes of data observed in the past few years.

- The benefits to model forecasts from all the observation types used in the Mesoscale model are reassuring because they confirm that all the data types are useful on (high-impact) occasions. The impacts should be taken into account when redesigning observation networks over and around the UK.

- It is anticipated that the signals from the results presented in this investigation will also apply when using the 'new dynamics' version of the Mesoscale model. Any differences in magnitude of impact are most likely to be found with MOPS, but this has yet to be tested.

- The increase in the relative benefit of aircraft data is consistent with the increased frequency of AMDAR reports. It is recommended that consideration be given to whether the current thinning of these data in the operational procedure is optimal or if

"super-obbing" should be introduced.

- An enhancement of the wind profiler network has already been recommended (at a time when the radiosonde network is reducing) by Bruce Macpherson (personal communication), partly based on evidence of the benefit of wind profile information from this study and Study A.

- It is recommended to study a few cases where observations have caused a marked degradation to the forecast, to determine if any one observation type (or its assimilation) is responsible for the small number of negative data impacts.

Acknowledgements

The author would like to thank Bruce Macpherson, Michael Bader and Andrew Lorenc for their guidance and Stephen Higginson for his help with the retrieval of archived Nimrod rainfall analyses.

References

Anderson, S.R., 1998: The impact of observations on Limited Area Model forecasts during the FASTEX campaign. Met Office Forecasting Research Tech. Report No. 251.

Anderson, S.R., R.J. Graham and M.J. Bader, 2000 ("Study A"): The impact of observations on Mesoscale Model forecasts of three-hourly rainfall accumulations. Meteorol. Appl., **7**, 193-200.

Graham, R.J., S.R. Anderson and M.J. Bader, 2000: The relative utility of current observation systems to global-scale NWP forecasts. Q. J. R. Meteorol. Soc., **126**, 2435-2460.

Kitchen, M., R. Brown and A.G. Davies, 1994: Real-time correction of weather radar data for the effects of bright band, range and orographic growth in widespread precipitation. Q. J. R. Meteorol. Soc., **120**, 1231-1254.

Jones, C. D. and B. Macpherson, 1997: A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. Meteorol. Appl., **4**, 269-277.

Lorenc, A.C., R.S. Bell, B. Macpherson, 1991: The Meteorological Office analysis correction scheme. Q. J. R. Meteorol. Soc., **177**, 59-89.

Macpherson, B., B.J. Wright, W.H. Hand and A.J. Maycock, 1996: The impact of MOPS moisture data in the UK Meteorological Office Mesoscale data assimilation scheme. Mon. Wea. Rev., **124**, 1746-1766.

Macpherson, B., P. Andrews, S. Harcourt, B. Ingleby, B. Chalcraft, A. Maycock, R.

Renshaw, C. Parrett, S. Anderson, M. Sharpe, D. Harrison and M. Gibson, 2002: The operational Mesoscale data assimilation system 1999-2001: Implementation of 3DVar and later upgrades. Met Office Forecasting Research Tech. Report No. 374.

Schaefer, J.T., 1990: The critical success index as an indicator of warning skill. Weather and Forecasting, **5**, 570-575.

Wilson, C. and I. Culverwell, 1997: New Mesoscale model configuration for the Cray T3E: Report on tests to identify "best domain and resolution". Met Office Forecasting Research Tech. Report No. 228.

Appendix 1: Further details of case selection technique

Cases for detailed study were selected using the following sequence of criteria:

(i) For this study, only cases in which the observed rainfall was widespread were considered for selection. The criterion used for 'widespread rain' was that the analysed 3-hour rainfall accumulations must exceed 0.5mm at more than 200 Mesoscale model grid points.

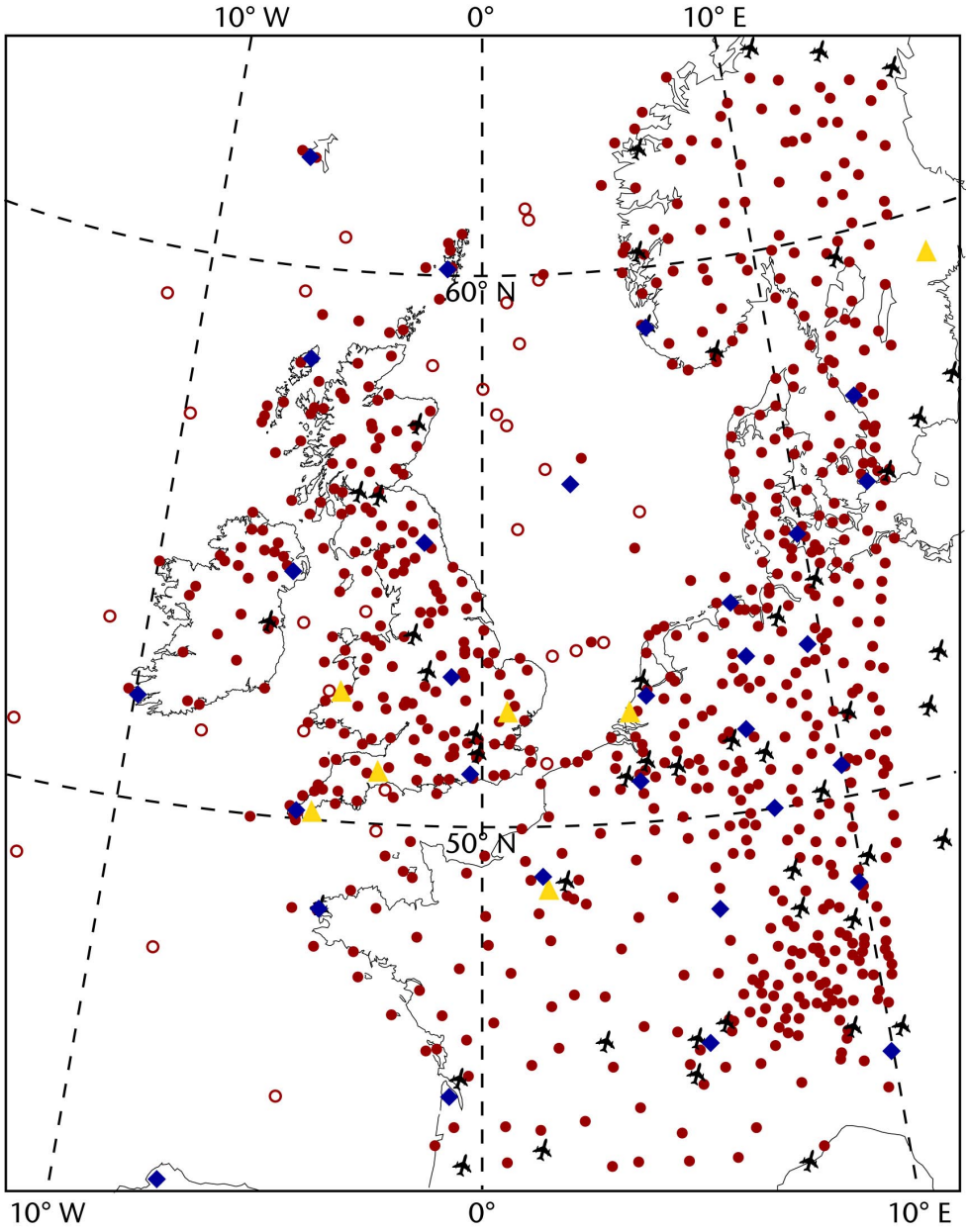
(ii) Secondly, objective methods were used to prepare a 'shortlist' of possible cases. We looked for a minimum acceptable improvement between consecutive forecasts, as measured by differences in the Equitable Threat Score (ETS - see Section 3) for rainfall accumulation. Experience showed that significant local improvements in the rainfall forecast were usually associated with improvements in the ETS of a factor of 2.0 or more (measured over the entire radar area for a threshold of 0.5mm/3h). This factor was therefore set as a criterion for defining a 'marked improvement'. Further criteria were also defined:

a. The false alarm rate must improve (decrease) by at least 5% (for at least one threshold),
b. The hit rate must improve (increase) by at least 5% (for at least one threshold),
c. The 0.1mm threshold hit rate for the later (shorter range) forecast must be at least 60%,
d. The ETS for the later forecast must be at least 0.1 (for at least one threshold), and
e. The ETS must improve (increase) by a ratio of 2.0 or by at least 0.15 (for at least one threshold).

(iii) Finally, subjective criteria based on examination of the rainfall distribution over the UK were applied. For example, only cases in which there were noticeable improvements in the positioning or accumulations associated with a well-formed rainband were considered. Many cases were discarded because the rainfall patterns, although 'widespread' by our definition in (i), were not organised into any recognisable structures.

N.B. The method of compiling the datasets for this case selection process has recently been updated. The suite runs automatically four times a day via a Suite Control System job, accumulates contingency tables on the hp system and archives the forecast and verification PP fields on the MASS archive. Further details can be acquired on request.

Appendix 2: Map showing locations of various observations within the Mesoscale model domain



- Upper-air station
- Wind profiler
- Airport served by AMDAR-equipped planes
- Land SYNOP station
- Moored buoy or rig

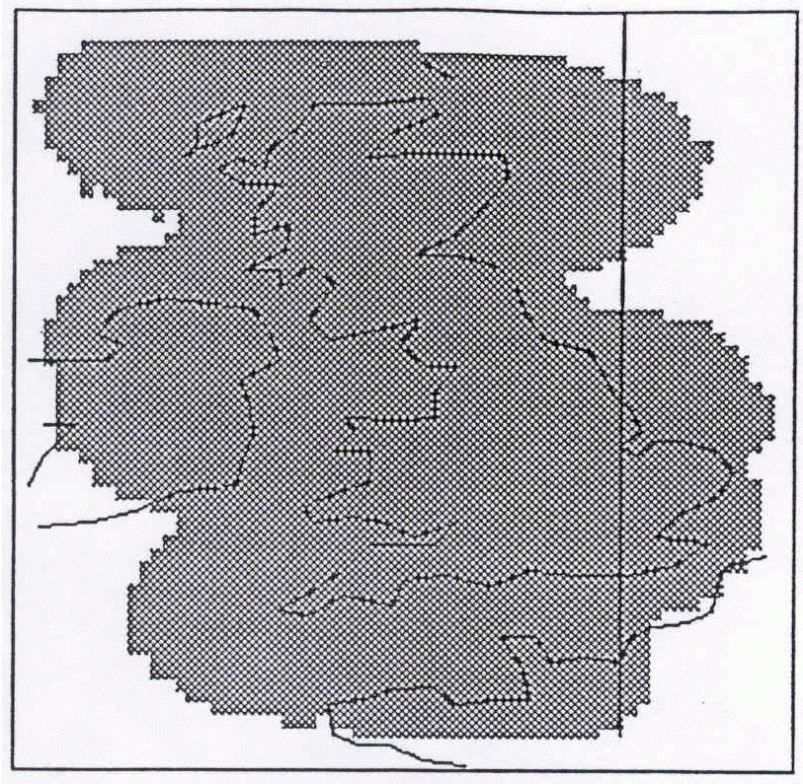


Fig. 1: Area covered (dark shading) by the UK weather radar network (at 100km range), used to generate the Nimrod accumulation analyses for forecast verification.

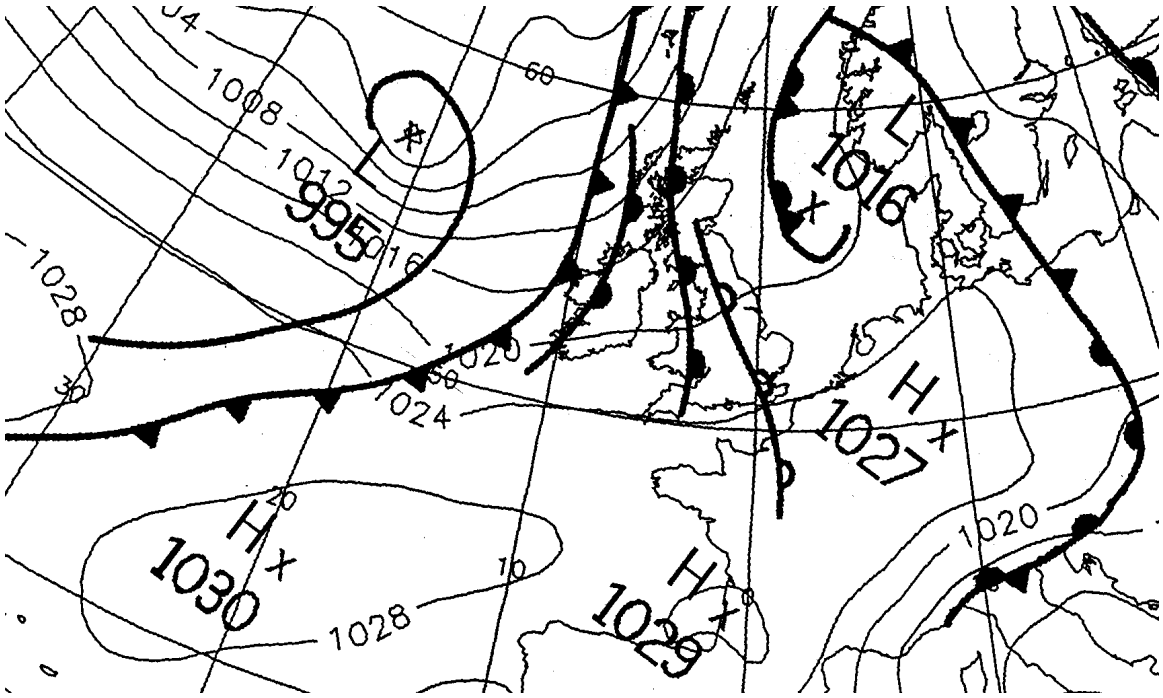


Fig. 2a: PMSL analysis for 06UTC 7th October 2000.

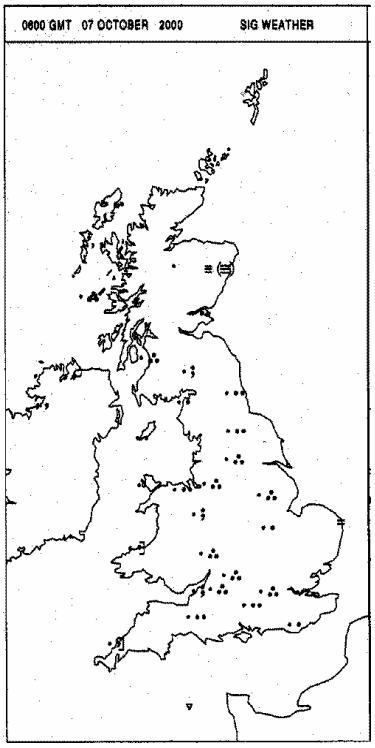


Fig. 2b: Significant weather chart for 06UTC 7th October 2000.

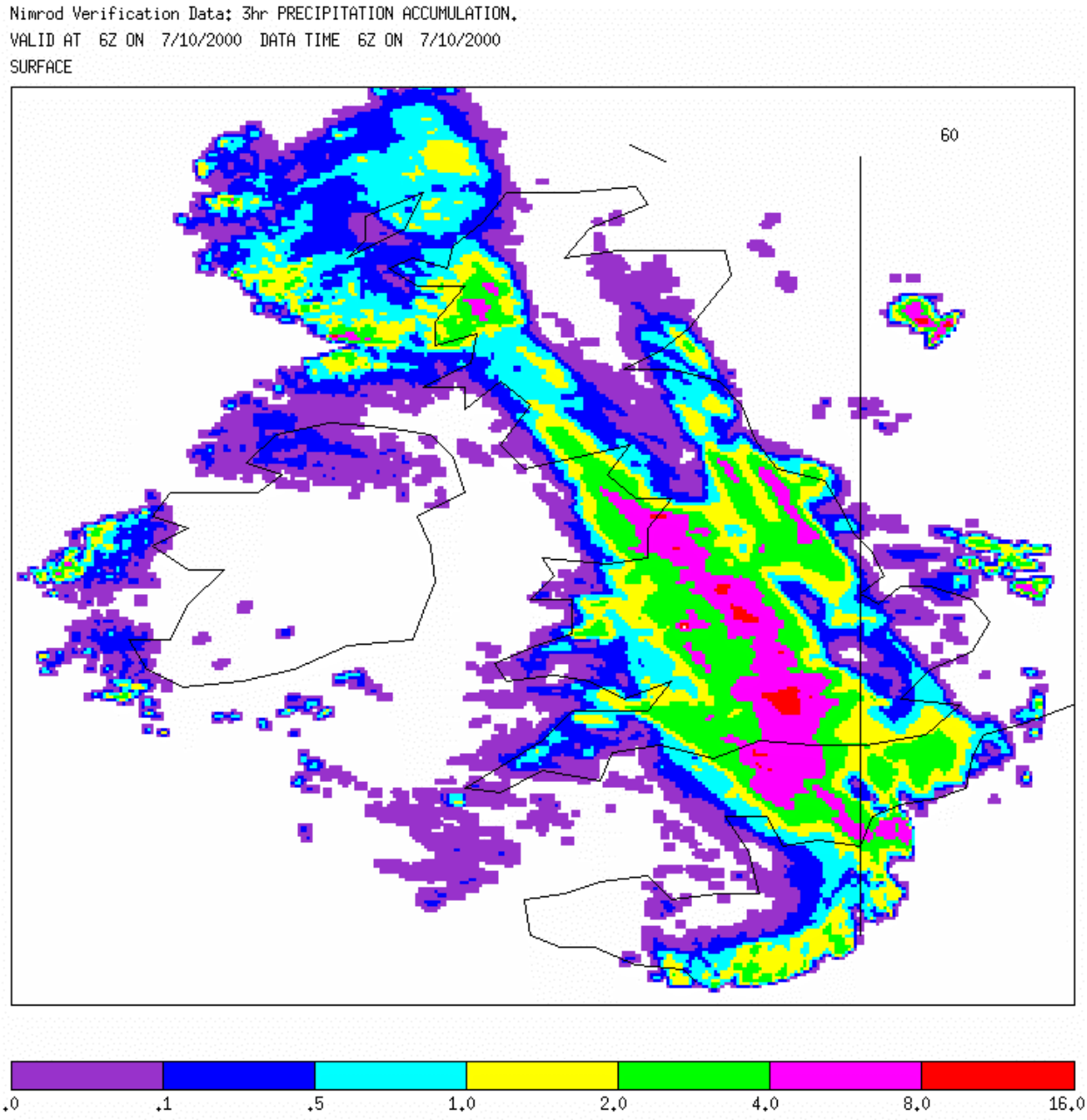


Fig. 3: Contoured precipitation accumulation field for the 3-hour period 03-06UTC 7th October 2000 (derived from Nimrod analyses). Units: mm.

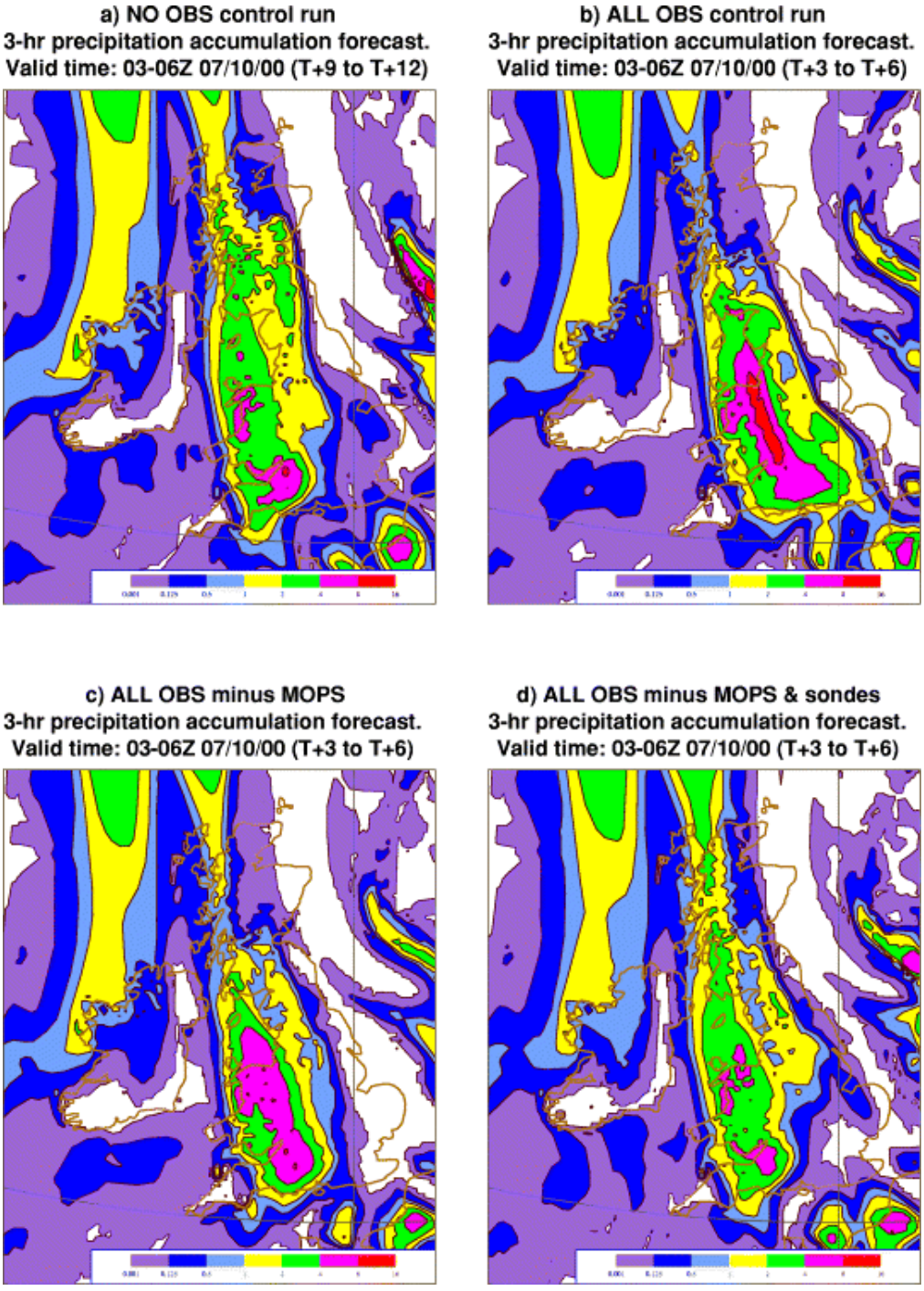


Fig. 4: Forecast precipitation accumulations for the 3-hour period 03-06UTC 7th October 2000 from **a)** NO OBS control run, **b)** ALL OBS control run, **c)** ALL OBS minus MOPS data denial run and **d)** ALL OBS minus

MOPS and Sondes data denial run. Units: mm. Same colour scale as in Fig. 3.

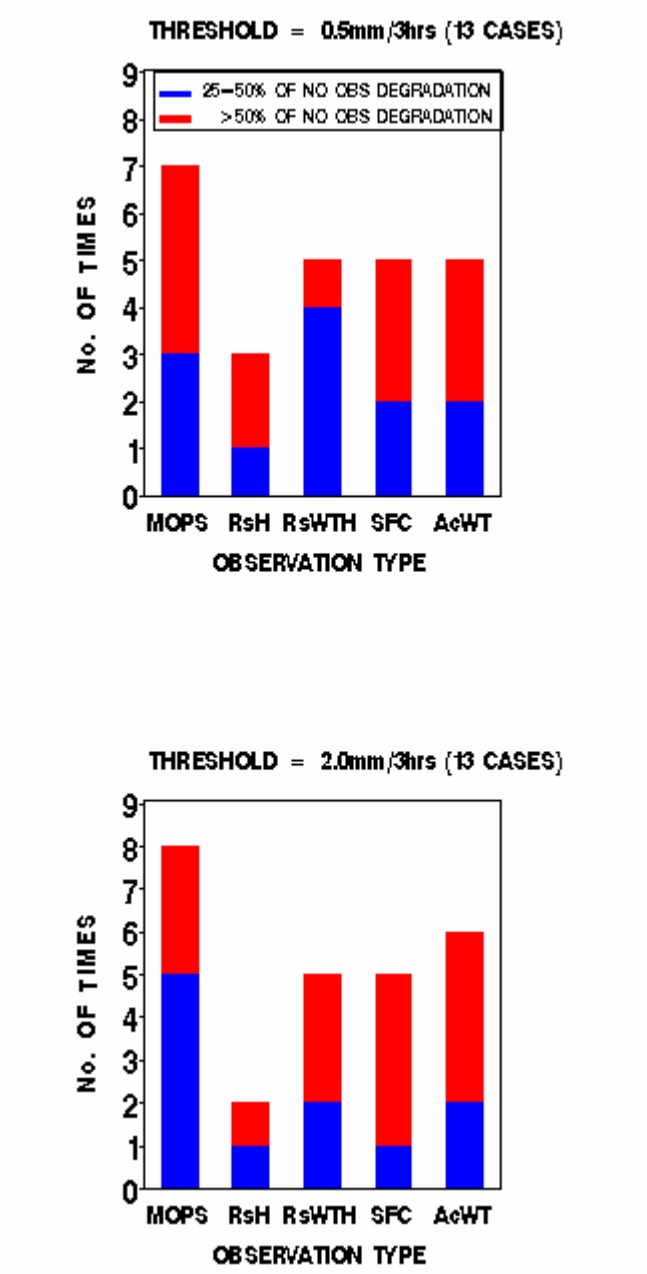


Fig. 5: Number of times removing an observation type delivered an individual degradation of 25% or more over 13 forecasts of precipitation accumulation above 0.5mm/3h (top plot) and above 2.0mm/3h (bottom plot). Blue bars show the number of times with degradation = 25-50%; red bars show number of times with degradation > 50%. MOPS = 3-D moisture field, RsH = radiosonde humidities, RsWTH = radiosonde winds, temperatures and humidities, SFC = surface pressure, wind, temperature and humidity, AcWT = aircraft winds and temperatures.

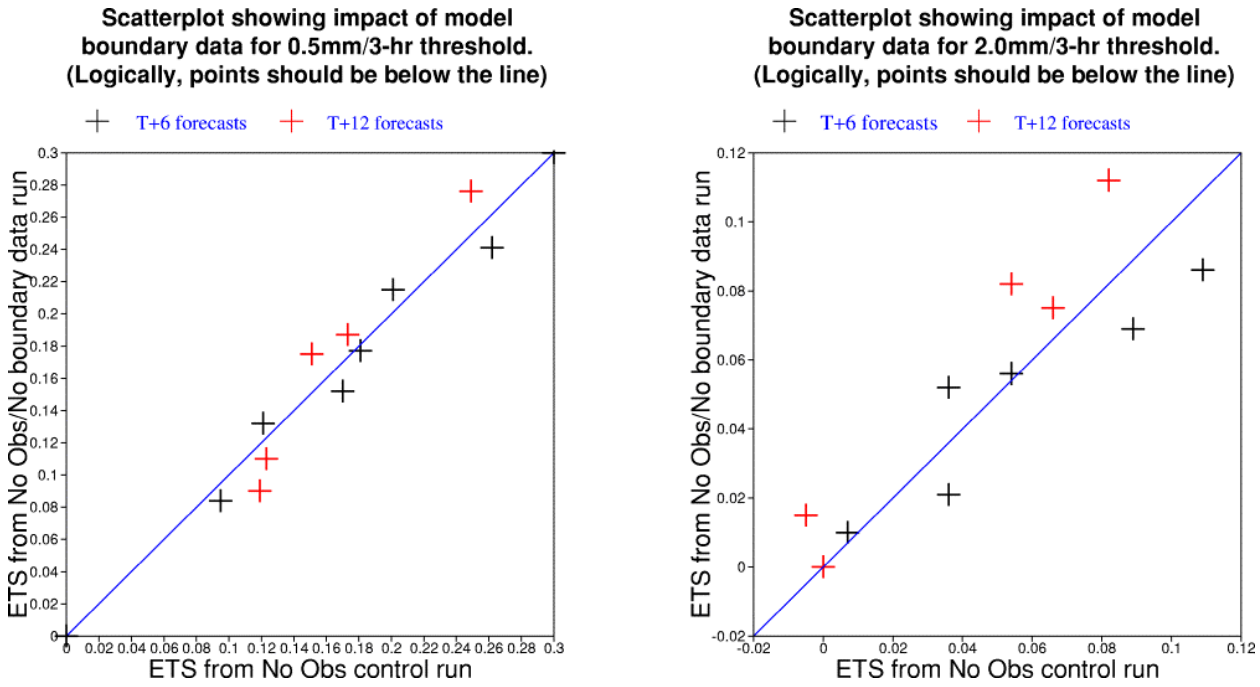


Fig. 6: Scatterplots showing ETS from the NO OBS control run vs ETS from the run using no observations or updated boundary data, for the first eleven cases, for precipitation accumulations above 0.5mm/3h (left plot) and above 2.0mm/3h (right plot). See Sub-section 5.2 for details.

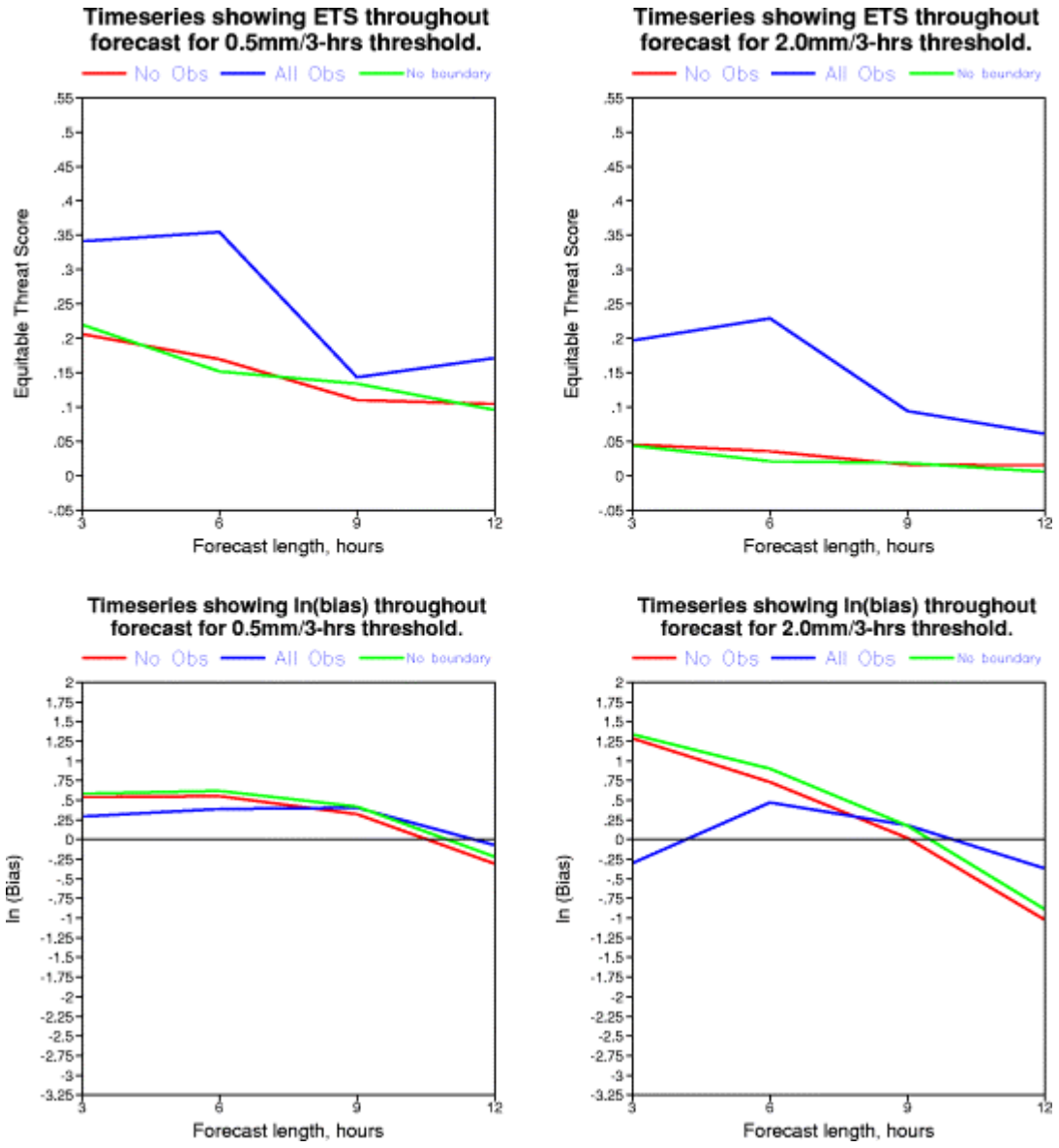


Fig. 7: Timeseries showing ETS and ln(bias) throughout the forecasts for Case 10. See individual captions and the text in Section 6 (vii) for details.